PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/04, C12N 15/00, 1/20, 9/00, A61K 38/43

A1

(11) International Publication Number: WO 97/39011

(43) International Publication Date: 23 October 1997 (23.10.97)

(21) International Application Number: PCT/US97/06551

(22) International Filing Date: 18 April 1997 (18.04.97)

(30) Priority Data:
9608000.7 18 April 1996 (18.04.96) GB

(71) Applicants (*for all designated States except US*): SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US). SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(72) Inventor; and
(75) Inventor/Applicant (*for US only*): LAWLOR, Elizabeth, Jane [GB/US]; 260 Lapp Road, Malvern, PA 19355 (US).

(74) Agents: GIMMI, Edward, R. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: STREPTOCOCCUS PNEUMONIAE ISOLEUCYL tRNA SYNTHETASE

(57) Abstract

The invention provides ileS polypeptides and DNA (RNA) encoding ileS polypeptides and methods for producing such polypeptides by recombinant techniques. Also provided are methods for utilizing ileS polypeptides to screen for antibacterial compounds.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece		Republic of Macedonia	**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's	**NZ**	New Zealand		
CM	Cameroon		Republic of Korea	**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

"Streptococcus pneumoniae Isoleucyl tRNA Synthetase".

RELATED APPLICATIONS

This application claims the benefit of UK application number 9608000.7, filed April 18, 1996.

FIELD OF THE INVENTION

This invention relates to newly identified polynucleotides and polypeptides, and their production and uses, as well as their variants, agonists and antagonists, and their uses. In particular, in these and in other regards, the invention relates to novel polynucleotides and polypeptides of the isoleucyl tRNA synthetase family, hereinafter referred to as "ileS".

BACKGROUND OF THE INVENTION

The Streptococci make up a medically important genera of microbes known to cause several types of disease in humans, including, for example, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid. Since its isolation more than 100 years ago, *Streptococcus pneumoniae* has been one of the more intensively studied microbes. For example, much of our early understanding that DNA is, in fact, the genetic material was predicated on the work of Griffith and of Avery, Macleod and McCarty using this microbe. Despite the vast amount of research with *S. pneumoniae*, many questions concerning the virulence of this microbe remain. It is particularly preferred to employ Streptococcal genes and gene products as targets for the development of antibiotics.

The frequency of *Streptococcus pneumoniae* infections has risen dramatically in the past 20 years. This has been attributed to the emergence of multiply antibiotic resistant strains and an increasing population of people with weakened immune systems. It is no longer uncommon to isolate *Streptococcus pneumoniae* strains which are resistant to some or all of the standard antibiotics. This has created a demand for both new anti-microbial agents and diagnostic tests for this organism.

The t-RNA synthetases have a primary role in protein synthesis according to the following scheme:

$$\text{Enzyme} + \text{ATP} + \text{AA} \Leftrightarrow \text{Enzyme.AA-AMP} + \text{PPi}$$

$$\text{Enzyme.AA-AMP} + \text{t-RNA} \Leftrightarrow \text{Enzyme} + \text{AMP} + \text{AA-t-RNA}$$

in which AA is an amino acid.

Inhibition of this process leads to a reduction in the levels of charged t-RNA and this triggers a cascade of responses known as the stringent response, the result of which is the

induction of a state of dormancy in the organism. As such selective inhibitors of bacterial t-RNA synthetase have potential as antibacterial agents. One example of such is mupirocin which is a selective inhibitor of isoleucyl t-RNA synthetase. Other t-RNA synthetases are now being examined as possible anti-bacterial targets, this process being greatly assisted by the isolation of the synthetase.

Clearly, there is a need for factors, such as the novel compounds of the invention, that have a present benefit of being useful to screen compounds for antibiotic activity. Such factors are also useful to determine their role in pathogenesis of infection, dysfunction and disease. There is also a need for identification and characterization of such factors and their antagonists and agonists which can play a role in preventing, ameliorating or correcting infections, dysfunctions or diseases.

The polypeptides of the invention have amino acid sequence homology to a known *Staphylococcus aureus* isoleucyl tRNA synthetase protein.

SUMMARY OF THE INVENTION

It is an object of the invention to provide polypeptides that have been identified as novel ileS polypeptides by homology between the amino acid sequence set out in Table 1 [SEQ ID NO: 2] and a known amino acid sequence or sequences of other proteins such as *Staphylococcus aureus* isoleucyl tRNA synthetase protein.

It is a further object of the invention to provide polynucleotides that encode ileS polypeptides, particularly polynucleotides that encode the polypeptide herein designated ileS.

In a particularly preferred embodiment of the invention the polynucleotide comprises a region encoding ileS polypeptides comprising at least one sequence set out in Table 1 [SEQ ID NOS:1, 5, 7], or a variant thereof.

In another particularly preferred embodiment of the invention there is a novel ileS protein from *Streptococcus pneumoniae* comprising an amino acid sequence of Table 1 [SEQ ID NOS:2, 6], or a variant thereof.

In accordance with another aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Streptococcus pneumoniae* 0100993 strain contained in the deposited strain.

A further aspect of the invention there are provided isolated nucleic acid molecules encoding ileS, particularly *Streptococcus pneumoniae* ileS, including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically,

prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of ileS and polypeptides encoded thereby.

Another aspect of the invention there are provided novel polypeptides of *Streptococcus pneumoniae* referred to herein as ileS as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of ileS polypeptide encoded by naturally occurring alleles of the ileS gene.

In a preferred embodiment of the invention there are provided methods for producing the aforementioned ileS polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing ileS expression, treating disease, for example, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid, assaying genetic variation, and administering a ileS polypeptide or polynucleotide to an organism to raise an immunological response against a bacteria, especially a *Streptococcus pneumoniae* bacteria.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided polynucleotides that hybridize to ileS polynucleotide sequences, particularly under stringent conditions.

In certain preferred embodiments of the invention there are provided antibodies against ileS polypeptides.

In other embodiments of the invention there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide

to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypetide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided ileS agonists and antagonists, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention there are provided compositions comprising a ileS polynucleotide or a ileS polypeptide for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math.*, *48:* 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity

are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research 12(1)*: 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol. 215*: 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al.*, *J. Mol. Biol. 215*: 403-410 (1990). As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence of SEQ ID NO: 1 it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence of SEQ ID NO: 1. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5 or 3 terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously , by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence of SEQ ID NO:2 an/or 6 is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of SEQ ID NO: 2. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by

natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol. 182*:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

"Variant(s)" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions,

deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

DESCRIPTION OF THE INVENTION

The invention relates to novel ileS polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of a novel ileS of *Streptococcus pneumoniae*, which is related by amino acid sequence homology to *Staphylococcus aureus* isoleucyl tRNA synthetase polypeptide. The invention relates especially to ileS having the nucleotide and amino acid sequences set out in Table 1 [SEQ ID NO: 1] and Table 1 [SEQ ID NO: 2] respectively, and to the ileS nucleotide sequences of the DNA in the deposited strain and amino acid sequences encoded thereby.

TABLE 1

ileS Polynucleotide and Polypeptide Sequences

(A) Sequences from *Streptococcus pneumoniae* ileS polynucleotide sequence.

Frament 1 [SEQ ID NO:1]

```
5'-1 ATGAAACTCA AAGACACCCT TAATCTTGGG AAAACTGAAT TCCCAATGCG

  51 TGCAGGCCTT CCTACCAAAG AGCCAGTTTG GCAAAAGGAA TGGGAAGATG

 101 CAAAACTTTA TCAACGTCGT CAAGAATTGA ACCAAGGAAA ACCTCATTTC

 151 ACCTTGCATG ATGGCCCTCC ATACGCTAAC GGAAATATCC ACGTTGGACA

 201 TGCTATGAAC AAGATTTCAA AAGATATCAT TGTTCGTTCT AAGTCTATGT

 251 CAGGATTTTA CGCGCCATTT ATTCCTGGTT GGGATACTCA TGGTCTGCCA
```

301 ATCGAGCAAG TCTTGTCAAA ACAAGGTGTC AAACGTAAAG AAATGGACTT

351 GGTTGAGTAC TTGAAACTTT GCCGTGAGTA CGCTCTTTCT CAAGTAGATA

401 AACAACGTGA AGATTTTAAA CGTTTGGGTG TTTCTGGTGA CTGGGAAAAT

451 CCATATGTGA CCTTGACTCC TGACTATGAA GCAGCTCAAA TTCGTGTATT

501 TGGTGAGATG GCTAATAAGG GTTATATCTA CCGTGGTGCC AAGCCAGTTT

551 ACTGGTCATG GTCATCTGAG TCAGCCCTTG CTGAAGCAGA GATTGAATAC

601 CATGACTTGG TTTCAACTTC CCTTTACTAT GCCAACAAGG TAAAAGATGG

651 CAAAGGAGTT CTAGATACAG ATACTTATAT CGTTGTCTGG ACAACGACTC

701 CATTTACCAT CACAGCTTCT CGTGGTTTGA CGGTTGGTGC AGATATTGAT

751 TACGTTTTGG TTCAACCTGC TGGTGAAGCT CGTAAGTTTG TCGTTGCTGC

801 TGAATTATTG ACTAG-3'

Fragent 2 [SEQ ID NO:5]

5'-1 TTGTCTGAGA AATTTGGCTG GGCTGATGTT CAAGTTTTGG AAACTTACCG

51 TGGCCAAGAA CTTAACCACA TCGTAACAGA ACACCCATGG GATACAGCTG

101 TAGAAGAGTT GGTAATTCTT GGTGACCACG TTACGACTGA CTCTGGTACA

151 GGTATTGTCC ATACAGCCCC TGGTTTTGGT GAGGACGACT ACAATGTTGG

201 TATTGCTAAT AATCTTGAAG TCGCAGTGAC TGTTGATGAA CGTGGTATCA

251 TGATGAAGAA TGCTGGTCCT GAGTTTGAAG GTCAATTCTA TGAAAAGGTA

301 GTTCCAACTG TTATTGAAAA ACTTGGTAAC CTCCTTCTTG CCCAAGAAGA

351 AATCTCTCAC TCATATCCAT TTGACTGGCG TACTAAGAAA CCAATCATCT

401 GGCGTGCAGT TCCACAATGG TTTGCCTCAG TTTCTAAATT CCGTCAAGAA

451 ATCTTGGACG AAATTGAAAA AGTGAAATTC CACTCAGAAT GGGGTAAAGT

501 CCGTCTTTAC AATATGATCC GTGACCGTGG TGACTGGGTT ATCTCTCGTC

551 AACGTGCTTG GGGTGTTCCA CTTCCAATCT TCTATGCAGA AGACGGTACA

601 GCTATCATGG TAGCTGAAAC GATTGAACAC GTAGCTCAAC TTTTTGAAGA

651 ACATGGTTCA AGCATTTGGT GGGAACGTGA TGCCAAAGAT CTCTTGCCAG

701 AAGGATTTAC TCATCCAGGT TCACCAAACG GCGAGTTCAA AAAAGAAACT

751 GATATCATGG ACGTTTGGTT TGACTCAGGT TCATCATGGA ATGGAGTGGT

801 GGTAAACCGT CCTGAATTGA CTTACCCAGC CGACCTTTAC CTAGAAGGTT

851 CTGACCAATA CCGTGGTTGG TTTAACTCAT CACTTATCAC ATCTGTTGCC

901 AACCATGGCG TAGCACCTTA CAAACAAATC TTGTCACAAG GTTTTGCCCT

951 TGATGGTAAA GGTGAGAAGA TGTCTAAATC TCTTGGAAAT ACCATTGCTC

1001 CAAGCGATGT TGAAAAACAA TTCGGTGCTG AAATCTTGCG TCTCTGGGTA

1051 ACAAGTGTTG ACTCAAGCAA TGACGTGCGT ATCTCTATGG ATATTTTGAG

1101 CCAAGTTTCT GAAACTTACC GTAAGATTCG TAACACTCTT CGTTTCTTGA

1151 TTGCCAATAC ATCTGACTTT AACCCAGCTC AAGATACAGT CGCTTACGAT

1201 GAGCTTCGTT CAGTTGATAA GTACATGACG ATTCGCTTTA ACCAGCTTGT

1251 CAAGACCATT CGTGATGCCT ATGCAGACTT TGAATTCTTG ACGATCTACA

1301 AGGCCTTGGT GAACTTTATC AACGTTGACT TGTCAGCCTT CTACCTTGAT

1351 TTTGCCAAAG ATGTTGTTTA CATTGAAGGT GCCAAATCAC TGGAACGCCG

1401 TCAAATGCAG ACTGTCTTCT ATGACATTCT TGTCAAAATC ACCAAACTCT

1451 TGACACCAAT CCTTCCTCAC ACTGCGGAAG AAATTTGGTC ATATCTTGAG

1501 TTTGAAACAG AAGACTTCGT CCAATTGTCA GAATTACCAG AGGCTCAAAC

1551 TTTTGCTAAT CAAGAAGAAA TCTTGGATAC ATGGGCAGCC TTCATGGACT

1601 TCCGTGGACA AGCTCAAAAA GCCTTGGAAG AAGCTCGTAA TGCAAAAGTA

1651 ATCGGTAAAT CACTTGAAGC ACACTTGACA GTTTATCCAA ACGAAGTTGT

1701 GAAAACTCTA CTCGAAGCAG TAAACAGCAA TGTGGCTCAA CTTTTGATCG

1751 TGTCAGACTT GACCATCGCA GAAGGACCAG CTCCAGAAGC TGCCCTTAGC

1801 TTCGAAGATG TAGCCTTCAC AGTTGAACGC GCTGCAGGTG AAGTATGTGA

1851 CCGTTGCCGT CGTATTGACC CAACAACAGC AGAACGTAGC TACCAGGCAG

1901 TTATCTGTGA CCACTGTGCA AGCATCGTAG AAGAAAACTT TGCGGAAGCA

1951 GTCGCAGAAG GATTTGAAGA GAAATAA-3'

(B) ileS polypeptide sequence deduced from the polynucleotide sequence of SEQ ID NO:1 [SEQ ID NO:2].

NH_2-1 MKLKDTLNLG KTEFPMRAGL PTKEPVWQKE WEDAKLYQRR QELNQGKPHF

51 TLHDGPPYAN GNIHVGHAMN KISKDIIVRS KSMSGFYAPF IPGWDTHGLP

101 IEQVLSKQGV KRKEMDLVEY LKLCREYALS QVDKQREDFK RLGVSGDWEN

151 PYVTLTPDYE AAQIRVFGEM ANKGYIYRGA KPVYWSWSSE SALAEAEIEY

201 HDLVSTSLYY ANKVKDGKGV LDTDTYIVVW TTTPFTITAS RGLTVGADID

251 YVLVQPAGEA RKFVVAAELL T-COOH

ileS polypeptide sequence deduced from the polynucleotide sequence of SEQ ID NO:5 [SEQ ID NO:6].

NH_2-1 LSEKFGWADV QVLETYRGQE LNHIVTEHPW DTAVEELVIL GDHVTTDSGT

51 GIVHTAPGFG EDDYNVGIAN NLEVAVTVDE RGIMMKNAGP EFEGQFYEKV

101 VPTVIEKLGN LLLAQEEISH SYPFDWRTKK PIIWRAVPQW FASVSKFRQE

151 ILDEIEKVKF HSEWGKVRLY NMIRDRGDWV ISRQRAWGVP LPIFYAEDGT

```
201  AIMVAETIEH VAQLFEEHGS SIWWERDAKD LLPEGFTHPG SPNGEFKKET
251  DIMDVWFDSG SSWNGVVVNR PELTYPADLY LEGSDQYRGW FNSSLITSVA
301  NHGVAPYKQI LSQGFALDGK GEKMSKSLGN TIAPSDVEKQ FGAEILRLWV
351  TSVDSSNDVR ISMDILSQVS ETYRKIRNTL RFLIANTSDF NPAQDTVAYD
401  ELRSVDKYMT IRFNQLVKTI RDAYADFEFL TIYKALVNFI NVDLSAFYLD
451  FAKDVVYIEG AKSLERRQMQ TVFYDILVKI TKLLTPILPH TAEEIWSYLE
501  FETEDFVQLS ELPEAQTFAN QEEILDTWAA FMDFRGQAQK ALEEARNAKV
551  IGKSLEAHLT VYPNEVVKTL LEAVNSNVAQ LLIVSDLTIA EGPAPEAALS
601  FEDVAFTVER AAGEVCDRCR RIDPTTAERS YQAVICDHCA SIVEENFAEA
651  VAEGFEEK-COOH
```

(C) Polynucleotide sequence embodiments.

Fragent 1 [SEQ ID NO:1]

$X-(R_1)_n-1$ ATGAAACTCA AAGACACCCT TAATCTTGGG AAAACTGAAT TCCCAATGCG

```
 51  TGCAGGCCTT CCTACCAAAG AGCCAGTTTG GCAAAAGGAA TGGGAAGATG
101  CAAAACTTTA TCAACGTCGT CAAGAATTGA ACCAAGGAAA ACCTCATTTC
151  ACCTTGCATG ATGGCCCTCC ATACGCTAAC GGAAATATCC ACGTTGGACA
201  TGCTATGAAC AAGATTTCAA AAGATATCAT TGTTCGTTCT AAGTCTATGT
251  CAGGATTTTA CGCGCCATTT ATTCCTGGTT GGGATACTCA TGGTCTGCCA
301  ATCGAGCAAG TCTTGTCAAA ACAAGGTGTC AAACGTAAAG AAATGGACTT
351  GGTTGAGTAC TTGAAACTTT GCCGTGAGTA CGCTCTTTCT CAAGTAGATA
401  AACAACGTGA AGATTTTAAA CGTTTGGGTG TTTCTGGTGA CTGGGAAAAT
451  CCATATGTGA CCTTGACTCC TGACTATGAA GCAGCTCAAA TTCGTGTATT
```

501 TGGTGAGATG GCTAATAAGG GTTATATCTA CCGTGGTGCC AAGCCAGTTT

551 ACTGGTCATG GTCATCTGAG TCAGCCCTTG CTGAAGCAGA GATTGAATAC

601 CATGACTTGG TTTCAACTTC CCTTTACTAT GCCAACAAGG TAAAAGATGG

651 CAAAGGAGTT CTAGATACAG ATACTTATAT CGTTGTCTGG ACAACGACTC

701 CATTTACCAT CACAGCTTCT CGTGGTTTGA CGGTTGGTGC AGATATTGAT

751 TACGTTTTGG TTCAACCTGC TGGTGAAGCT CGTAAGTTTG TCGTTGCTGC

801 TGAATTATTG ACTAG-$(R_2)_n$-Y

Fragent 2 [SEQ ID NO:5]

X-$(R_1)_n$-1 TTGTCTGAGA AATTTGGCTG GGCTGATGTT CAAGTTTTGG AAACTTACCG

51 TGGCCAAGAA CTTAACCACA TCGTAACAGA ACACCCATGG GATACAGCTG

101 TAGAAGAGTT GGTAATTCTT GGTGACCACG TTACGACTGA CTCTGGTACA

151 GGTATTGTCC ATACAGCCCC TGGTTTTGGT GAGGACGACT ACAATGTTGG

201 TATTGCTAAT AATCTTGAAG TCGCAGTGAC TGTTGATGAA CGTGGTATCA

251 TGATGAAGAA TGCTGGTCCT GAGTTTGAAG GTCAATTCTA TGAAAAGGTA

301 GTTCCAACTG TTATTGAAAA ACTTGGTAAC CTCCTTCTTG CCCAAGAAGA

351 AATCTCTCAC TCATATCCAT TTGACTGGCG TACTAAGAAA CCAATCATCT

401 GGCGTGCAGT TCCACAATGG TTTGCCTCAG TTTCTAAATT CCGTCAAGAA

451 ATCTTGGACG AAATTGAAAA AGTGAAATTC CACTCAGAAT GGGGTAAAGT

501 CCGTCTTTAC AATATGATCC GTGACCGTGG TGACTGGGTT ATCTCTCGTC

551 AACGTGCTTG GGGTGTTCCA CTTCCAATCT TCTATGCAGA AGACGGTACA

601 GCTATCATGG TAGCTGAAAC GATTGAACAC GTAGCTCAAC TTTTTGAAGA

651 ACATGGTTCA AGCATTTGGT GGGAACGTGA TGCCAAAGAT CTCTTGCCAG

701 AAGGATTTAC TCATCCAGGT TCACCAAACG GCGAGTTCAA AAAAGAAACT

751 GATATCATGG ACGTTTGGTT TGACTCAGGT TCATCATGGA ATGGAGTGGT

801 GGTAAACCGT CCTGAATTGA CTTACCCAGC CGACCTTTAC CTAGAAGGTT

851 CTGACCAATA CCGTGGTTGG TTTAACTCAT CACTTATCAC ATCTGTTGCC

901 AACCATGGCG TAGCACCTTA CAAACAAATC TTGTCACAAG GTTTTGCCCT

951 TGATGGTAAA GGTGAGAAGA TGTCTAAATC TCTTGGAAAT ACCATTGCTC

1001 CAAGCGATGT TGAAAAACAA TTCGGTGCTG AAATCTTGCG TCTCTGGGTA

1051 ACAAGTGTTG ACTCAAGCAA TGACGTGCGT ATCTCTATGG ATATTTTGAG

1101 CCAAGTTTCT GAAACTTACC GTAAGATTCG TAACACTCTT CGTTTCTTGA

1151 TTGCCAATAC ATCTGACTTT AACCCAGCTC AAGATACAGT CGCTTACGAT

1201 GAGCTTCGTT CAGTTGATAA GTACATGACG ATTCGCTTTA ACCAGCTTGT

1251 CAAGACCATT CGTGATGCCT ATGCAGACTT TGAATTCTTG ACGATCTACA

1301 AGGCCTTGGT GAACTTTATC AACGTTGACT TGTCAGCCTT CTACCTTGAT

1351 TTTGCCAAAG ATGTTGTTTA CATTGAAGGT GCCAAATCAC TGGAACGCCG

1401 TCAAATGCAG ACTGTCTTCT ATGACATTCT TGTCAAAATC ACCAAACTCT

1451 TGACACCAAT CCTTCCTCAC ACTGCGGAAG AAATTTGGTC ATATCTTGAG

1501 TTTGAAACAG AAGACTTCGT CCAATTGTCA GAATTACCAG AGGCTCAAAC

1551 TTTTGCTAAT CAAGAAGAAA TCTTGGATAC ATGGGCAGCC TTCATGGACT

1601 TCCGTGGACA AGCTCAAAAA GCCTTGGAAG AAGCTCGTAA TGCAAAAGTA

1651 ATCGGTAAAT CACTTGAAGC ACACTTGACA GTTTATCCAA ACGAAGTTGT

1701 GAAAACTCTA CTCGAAGCAG TAAACAGCAA TGTGGCTCAA CTTTTGATCG

1751 TGTCAGACTT GACCATCGCA GAAGGACCAG CTCCAGAAGC TGCCCTTAGC

1801 TTCGAAGATG TAGCCTTCAC AGTTGAACGC GCTGCAGGTG AAGTATGTGA

1851 CCGTTGCCGT CGTATTGACC CAACAACAGC AGAACGTAGC TACCAGGCAG

1901 TTATCTGTGA CCACTGTGCA AGCATCGTAG AAGAAAACTT TGCGGAAGCA

1951 GTCGCAGAAG GATTTGAAGA GAAATAA-$(R_2)_n$-Y

(D) Polypeptide sequence embodiments [SEQ ID NO:2].

X-$(R_1)_n$-1 MKLKDTLNLG KTEFPMRAGL PTKEPVWQKE WEDAKLYQRR QELNQGKPHF

51 TLHDGPPYAN GNIHVGHAMN KISKDIIVRS KSMSGFYAPF IPGWDTHGLP

101 IEQVLSKQGV KRKEMDLVEY LKLCREYALS QVDKQREDFK RLGVSGDWEN

151 PYVTLTPDYE AAQIRVFGEM ANKGYIYRGA KPVYWSWSSE SALAEAEIEY

201 HDLVSTSLYY ANKVKDGKGV LDTDTYIVVW TTTPFTITAS RGLTVGADID

251 YVLVQPAGEA RKFVVAAELL T-$(R_2)_n$-Y

[SEQ ID NO:6]

X-$(R_1)_n$-1 LSEKFGWADV QVLETYRGQE LNHIVTEHPW DTAVEELVIL GDHVTTDSGT

51 GIVHTAPGFG EDDYNVGIAN NLEVAVTVDE RGIMMKNAGP EFEGQFYEKV

101 VPTVIEKLGN LLLAQEEISH SYPFDWRTKK PIIWRAVPQW FASVSKFRQE

151 ILDEIEKVKF HSEWGKVRLY NMIRDRGDWV ISRQRAWGVP LPIFYAEDGT

201 AIMVAETIEH VAQLFEEHGS SIWWERDAKD LLPEGFTHPG SPNGEFKKET

251 DIMDVWFDSG SSWNGVVVNR PELTYPADLY LEGSDQYRGW FNSSLITSVA

301 NHGVAPYKQI LSQGFALDGK GEKMSKSLGN TIAPSDVEKQ FGAEILRLWV

351 TSVDSSNDVR ISMDILSQVS ETYRKIRNTL RFLIANTSDF NPAQDTVAYD

401 ELRSVDKYMT IRFNQLVKTI RDAYADFEFL TIYKALVNFI NVDLSAFYLD

451 FAKDVVYIEG AKSLERRQMQ TVFYDILVKI TKLLTPILPH TAEEIWSYLE

501 FETEDFVQLS ELPEAQTFAN QEEILDTWAA FMDFRGQAQK ALEEARNAKV

551 IGKSLEAHLT VYPNEVVKTL LEAVNSNVAQ LLIVSDLTIA EGPAPEAALS

601 FEDVAFTVER AAGEVCDRCR RIDPTTAERS YQAVICDHCA SIVEENFAEA

651 VAEGFEEK-$(R_2)_n$-Y

(E) Polynucleotide sequence embodiment [SEQ ID NO:7].

5'-1 CAACTTTTTG AAGAACATGG TTCAAGCATT TGGTGGGAAC GTGATGCCAA

51 AGATCTCTTG CCAGAAGGAT TTACTCATCC AGGTTCACCA AACGGCGAGT

101 TCAAAAAAGA AACTGATATC ATGGACGTTT GGTTTGACTC AGGTTCATCA

151 TGGAATGGAG TGGTGGTAAA CCGTCCTGAA TTGACTTACC CAGCCGACCT

201 TTACCTAGAA GGTTCTGACC AATACCGTGG TTGGTTTAAC TCATCACTTA

251 TCACATCTGT TGCCAACCAT GGCGTAGCAC CTTACAAACA AATCTTGTCA

301 CAAGGTTTTG CCCTTGATGG TAAAGGTGAG AAGATGTCTA AATCTCTTGG

351 AAATACCATT GCTCCAAGCG ATGTTGAAAA ACAATTCGGG-3'

Deposited materials

A deposit containing a *Streptococcus pneumoniae* 0100993 strain has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (herein "NCIMB"), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 April 1996 and assigned deposit number 40794. The deposit was described as *Streptococcus pneumoniae* 0100993 on deposit. On 17 April 1996 a *Streptococcus pneumoniae* 0100993 DNA library in E. coli was similarly depositedwith the NCIMB and assigned deposit number 40800. The *Streptococcus pneumoniae* strain deposit is referred to herein as "the deposited strain" or as "the DNA of the deposited strain."

The deposited strain contains the full length ileS gene. The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

The deposit of the deposited strain has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposited strain is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

A license may be required to make, use or sell the deposited strain, and compounds derived therefrom, and no such license is hereby granted.

Polypeptides

The polypeptides of the invention include the polypeptides of Table 1 [SEQ ID NO:2, 6] (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of ileS, and also those which have at least 70% identity to a polypeptide of Table 1 [SEQ ID NO:2, 6] or the relevant portion, preferably at least 80% identity a polypeptide of Table 1 [SEQ ID NO:2, 6], and more preferably at least 90% similarity (more preferably at least 90% identity) to a polypeptide of Table 1 [SEQ ID NO:2, 6] and still more preferably at least 95% similarity (still more preferably at least 95% identity) to a polypeptide of Table 1 [SEQ ID NO:2, 6] and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

The invention also includes polypeptides of the formula set forth in Table 1 (D) wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_2 is any amino acid residue, and n is an integer between 1 and 1000. Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with ileS polypeptides fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of an amino acid sequence of Table 1 [SEQ ID NO:2, 6], or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the

invention in a host cell, particularly a *Streptococcus pneumoniae*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of ileS, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Streptococcus pneumoniae* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

Polynucleotides

Another aspect of the invention relates to isolated polynucleotides that encode the ileS polypeptide having a deduced amino acid sequence of Table 1 [SEQ ID NO:2, 6] and polynucleotides closely related thereto and variants thereof.

Using the information provided herein, such as a polynucleotide sequence set out in Table 1 [SEQ ID NOS:1, 5, 7], a polynucleotide of the invention encoding ileS polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Streptococcus pneumoniae* 0100993 cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a sequence given in Table 1 [SEQ ID NOS:1, 5, 7], typically a library of clones of chromosomal DNA of *Streptococcus pneumoniae* 0100993 in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both

directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, each polynucleotide set out in Table 1 [SEQ ID NOS:1, 5, 7] was discovered in a DNA library derived from *Streptococcus pneumoniae* 0100993.

Each DNA sequence set out in Table 1 [SEQ ID NOS:1, 5, 7] contains an open reading frame encoding a protein having about the number of amino acid residues set forth in Table 1 [SEQ ID NO:2, 6] with a deduced molecular weight that can be calculated using amino acid residue molecular weight values well known in the art. The start codon of the DNA in Table 1 is nucleotide number 1 and last codon that encodes an amino acid is number 815 for "Fragment 1" herein, and analogously 1 to 1974 for "Fragment 2" herein, the stop codon being the next codon following this last codon encoding an amino acid.

ileS of the invention is structurally related to other proteins of the isoleucyl tRNA synthetase family, as shown by the results of sequencing the DNA encoding ileS of the deposited strain. The protein exhibits greatest homology to *Staphylococcus aureus* isoleucyl tRNA synthetase protein among known proteins. ileS of Table 1 [SEQ ID NO:2] has about 60 - 54% identity over its entire length and about 74 - 71% similarity over its entire length with the amino acid sequence of *Staphylococcus aureus* isoleucyl tRNA synthetase polypeptide.

The invention provides polynucleotide sequences identical over its entire length to the coding sequence in Table 1 [SEQ ID NOS:1, 5, 7]. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and

described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA 86:* 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell 37:* 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

A preferred embodiment of the invention includes, for example, a polynucleotide comprising nucleotide 1 to 815 or 1 to 1974 set forth in SEQ ID NO:1 and SEQ ID NO:5 respectively of Table 1 each of which encodes ileS polypeptide.

The invention also includes polynucleotides of the formula set forth in Table 1 (C) wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal, R_1 and R_2 is any nucleic acid residue, and n is an integer between 1 and 1000. Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Streptococcus pneumoniae* ileS comprising an amino acid sequence set out in Table 1 [SEQ ID NO:2, 6]. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide comprising a deduced amino acid sequence of Table 1 [SEQ ID NO:2, 6]. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding ileS variants, that comprise the amino acid sequence of ileS polypeptide of Table 1 [SEQ ID NO:2, 6] in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of ileS.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical over their entire length to a polynucleotide encoding ileS polypeptide comprising an amino acid sequence set out in Table 1 [SEQ ID NO:2, 6], and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are

polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding ileS polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as a mature polypeptide comprising a polypeptide sequence encoded by the DNA of Table 1 [SEQ ID NOS:1, 5, 7].

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in SEQ ID NOS:1, 5, 7 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in SEQ ID NOS:1, 5, 7 or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization

probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding ileS and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the ileS gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of the ileS gene may be isolated by screening using the DNA sequence provided in SEQ ID NO: 1 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the sequences of SEQ ID NOS:1 and/or 2 and/or 5 and/or 6 and/or 7 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Vectors, host cells, expression

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic

elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention is also related to the use of the ileS polynucleotides of the invention for use as diagnostic reagents. Detection of ileS in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with an organism comprising the ileS gene may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA or cDNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of

a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled ileS polynucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, *e.g.*, Myers et al., *Science, 230:* 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, *e.g.*, Cotton et al., *Proc. Natl. Acad. Sci., USA, 85:* 4397-4401 (1985).

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to used RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding ileS can be used to identify and analyze mutations. Examples of representative primers are shown below in Table 2.

Table 2

Primers for amplification of ileS polynucleotides

SEQ ID NO	PRIMER SEQUENCE
3	5'-ATGAAACTCAAAGACACCCTTAAT-3'
4	5'-TTATTTCTCTTCAAATCCTTCTGCG-3'

The invention further provides these primers with 1, 2, 3 or 4 nucleotides removed from the 5' and/or the 3' end. These primers may be used for, among othe4r things, amplifying ileS DNA isolated from a sample derived from an individual. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably infections by *Streptococcus pneumoniae*, and most preferably otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid, comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having the sequence of Table 1 [SEQ ID NO: 1]. Increased or decreased expression of ileS polynucleotide can be measured using any on of the methods well known in the art for the quantation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of ileS protein compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a ileS protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunolglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature 256:* 495-497 (1975); Kozbor *et al., Immunology Today 4:* 72 (1983); Cole et al., pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also,

transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-ileS or from naive libraries (McCafferty, J. et al., (1990), Nature 348, 552-554; Marks, J. et al., (1992) Biotechnology *10*, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) Nature 352, 624-628).

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against ileS- polypeptide may be employed to treat infections, particularly bacterial infections and especially otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complimentarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody , for example as described in Jones, P. et al. (1986), Nature 321, 522-525 or Tempest et al.,(1991) Biotechnology 9, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff et al., Hum Mol Genet 1992, 1:363, Manthorpe et al., Hum. Gene Ther. 1963:4, 419), delivery of DNA complexed with specific protein carriers (Wu et al., J Biol Chem. 1989: 264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, PNAS, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda et al., Science 1989:243,375), particle bombardment (Tang et al., Nature 1992, 356:152, Eisenbraun et al., DNA Cell Biol 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., PNAS 1984:81,5849).

Antagonists and agonists - assays and molecules

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, *e.g.*, Coligan *et al.*, *Current Protocols in Immunology 1(2):* Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of ileS polypeptides or polynucleotides, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagoists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising ileS polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be a ileS agonist or antagonist. The ability of the candidate molecule to agonize or antagonize the ileS polypeptide is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of ileS polypeptide are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from

substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in ileS polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for ileS antagonists is a competitive assay that combines ileS and a potential antagonist with ileS-binding molecules, recombinant ileS binding molecules, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. ileS can be labeled, such as by radioactivity or a colorimetric compound, such that the number of ileS molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing ileS-induced activities, thereby preventing the action of ileS by excluding ileS from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of ileS.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block ileS protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al., Infect. Immun. 60*:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial ileS proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with ileS, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Streptococcus pneumoniae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of ileS, or a fragment or a variant thereof, for expressing ileS, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise.
Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a ileS or protein coded therefrom, wherein the composition comprises a recombinant ileS or protein coded therefrom comprising DNA which codes for and expresses an antigen of said ileS or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A ileS polypeptide or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Streptococcus pneumoniae* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Streptococcus pneumoniae* infection, in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking

adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation insotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain ileS protein, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions, kits and administration

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and

pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device.

Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Streptococcus pneumoniae* wound infections.

Many orthopaedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1μg/ml to 10mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

EXAMPLES

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1 Strain selection, Library Production and Sequencing

The polynucleotides having the DNA sequence given in SEQ ID NO:1, 5 and 7 were obtained from a library of clones of chromosomal DNA of *Streptococcus pneumoniae*

in *E. coli.* The sequencing data from two or more clones containing overlapping *Streptococcus pneumoniae* DNAs was used to construct the contiguous DNA sequence in SEQ ID NO:1, 5 and 7. Libraries may be prepared by routine methods, for example: Methods 1 and 2 below.

Total cellular DNA is isolated from *Streptococcus pneumoniae* 0100993 according to standard procedures and size-fractionated by either of two methods.

Method 1

Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 11kbp in size are rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library packaged by standard procedures and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

Method 2

Total cellular DNA is partially hydrolyzed with a one or a combination of restriction enzymes appropriate to generate a series of fragments for cloning into library vectors (e.g., RsaI, PalI, AluI, Bshl235I), and such fragments are size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

Example 2 ileS Characterization

The enzyme mediated incorporation of radiolabelled amino acid into tRNA may be measured by the aminoacylation method which measures amino acid-tRNA as trichloroacetic acid-precipitable radioactivity from radiolabelled amino acid in the presence of tRNA and ATP (Hughes J, Mellows G and Soughton S. 1980, FEBS Letters, 122:322-324). Thus inhibitors of isoleucyl tRNA synthetase can be detected by a reduction in the trichloroacetic acid precipitable radioactivity relative to the control. Alternatively the tRNA synthetase catalysed partial PPi/ATP exchange reaction which measures the formation of radiolabelled ATP from PPi can be used to detect isoleucyl tRNA synthetase inhibitors (Calender R & Berg P. 1966, Biochemistry, 5, 1681-1690).

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Lawlor, Elizabeth

(ii) TITLE OF THE INVENTION: Novel Compounds

(iii) NUMBER OF SEQUENCES: 7

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19046

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 18-APR-1997
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 9608000.7
(B) FILING DATE: 18-APR-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Gimmi, Edward R
(B) REGISTRATION NUMBER: 38,891
(C) REFERENCE/DOCKET NUMBER: P31455

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 610-270-4478
(B) TELEFAX: 610-270-5090
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 815 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATGAAACTCA AAGACACCCT TAATCTTGGG AAAACTGAAT TCCCAATGCG TGCAGGCCTT    60
CCTACCAAAG AGCCAGTTTG GCAAAAGGAA TGGGAAGATG CAAAACTTTA TCAACGTCGT   120
CAAGAATTGA ACCAAGGAAA ACCTCATTTC ACCTTGCATG ATGGCCCTCC ATACGCTAAC   180
GGAAATATCC ACGTTGGACA TGCTATGAAC AAGATTTCAA AAGATATCAT TGTTCGTTCT   240
AAGTCTATGT CAGGATTTTA CGCGCCATTT ATTCCTGGTT GGGATACTCA TGGTCTGCCA   300
ATCGAGCAAG TCTTGTCAAA ACAAGGTGTC AAACGTAAAG AAATGGACTT GGTTGAGTAC   360
TTGAAACTTT GCCGTGAGTA CGCTCTTTCT CAAGTAGATA AACAACGTGA AGATTTTAAA   420
CGTTTGGGTG TTTCTGGTGA CTGGGAAAAT CCATATGTGA CCTTGACTCC TGACTATGAA   480
GCAGCTCAAA TTCGTGTATT TGGTGAGATG GCTAATAAGG GTTATATCTA CCGTGGTGCC   540
AAGCCAGTTT ACTGGTCATG GTCATCTGAG TCAGCCCTTG CTGAAGCAGA GATTGAATAC   600
CATGACTTGG TTTCAACTTC CCTTTACTAT GCCAACAAGG TAAAAGATGG CAAAGGAGTT   660
CTAGATACAG ATACTTATAT CGTTGTCTGG ACAACGACTC CATTTACCAT CACAGCTTCT   720
CGTGGTTTGA CGGTTGGTGC AGATATTGAT TACGTTTTGG TTCAACCTGC TGGTGAAGCT   780
CGTAAGTTTG TCGTTGCTGC TGAATTATTG ACTAG                              815
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Lys Leu Lys Asp Thr Leu Asn Leu Gly Lys Thr Glu Phe Pro Met
1 5 10 15
Arg Ala Gly Leu Pro Thr Lys Glu Pro Val Trp Gln Lys Glu Trp Glu
20 25 30
Asp Ala Lys Leu Tyr Gln Arg Arg Gln Glu Leu Asn Gln Gly Lys Pro
35 40 45
His Phe Thr Leu His Asp Gly Pro Pro Tyr Ala Asn Gly Asn Ile His
50 55 60
Val Gly His Ala Met Asn Lys Ile Ser Lys Asp Ile Ile Val Arg Ser
65 70 75 80
Lys Ser Met Ser Gly Phe Tyr Ala Pro Phe Ile Pro Gly Trp Asp Thr
85 90 95
His Gly Leu Pro Ile Glu Gln Val Leu Ser Lys Gln Gly Val Lys Arg
100 105 110
Lys Glu Met Asp Leu Val Glu Tyr Leu Lys Leu Cys Arg Glu Tyr Ala
115 120 125
Leu Ser Gln Val Asp Lys Gln Arg Glu Asp Phe Lys Arg Leu Gly Val
130 135 140
Ser Gly Asp Trp Glu Asn Pro Tyr Val Thr Leu Thr Pro Asp Tyr Glu
145 150 155 160
Ala Ala Gln Ile Arg Val Phe Gly Glu Met Ala Asn Lys Gly Tyr Ile
165 170 175
Tyr Arg Gly Ala Lys Pro Val Tyr Trp Ser Trp Ser Ser Glu Ser Ala
180 185 190
Leu Ala Glu Ala Glu Ile Glu Tyr His Asp Leu Val Ser Thr Ser Leu
195 200 205
Tyr Tyr Ala Asn Lys Val Lys Asp Gly Lys Gly Val Leu Asp Thr Asp
210 215 220
Thr Tyr Ile Val Val Trp Thr Thr Thr Pro Phe Thr Ile Thr Ala Ser
225 230 235 240
Arg Gly Leu Thr Val Gly Ala Asp Ile Asp Tyr Val Leu Val Gln Pro
245 250 255
Ala Gly Glu Ala Arg Lys Phe Val Val Ala Ala Glu Leu Leu Thr
260 265 270

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGAAACTCA AAGACACCCT TAAT 24

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TTATTTCTCT TCAAATCCTT CTGCG 25

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1977 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTGTCTGAGA AATTTGGCTG GGCTGATGTT CAAGTTTTGG AAACTTACCG TGGCCAAGAA 60
CTTAACCACA TCGTAACAGA ACACCCATGG GATACAGCTG TAGAAGAGTT GGTAATTCTT 120
GGTGACCACG TTACGACTGA CTCTGGTACA GGTATTGTCC ATACAGCCCC TGGTTTTGGT 180
GAGGACGACT ACAATGTTGG TATTGCTAAT AATCTTGAAG TCGCAGTGAC TGTTGATGAA 240
CGTGGTATCA TGATGAAGAA TGCTGGTCCT GAGTTTGAAG GTCAATTCTA TGAAAAGGTA 300
GTTCCAACTG TTATTGAAAA ACTTGGTAAC CTCCTTCTTG CCCAAGAAGA AATCTCTCAC 360
TCATATCCAT TTGACTGGCG TACTAAGAAA CCAATCATCT GGCGTGCAGT TCCACAATGG 420
TTTGCCTCAG TTTCTAAATT CCGTCAAGAA ATCTTGGACG AAATTGAAAA AGTGAAATTC 480
CACTCAGAAT GGGGTAAAGT CCGTCTTTAC AATATGATCC GTGACCGTGG TGACTGGGTT 540
ATCTCTCGTC AACGTGCTTG GGGTGTTCCA CTTCCAATCT TCTATGCAGA AGACGGTACA 600
GCTATCATGG TAGCTGAAAC GATTGAACAC GTAGCTCAAC TTTTTGAAGA ACATGGTTCA 660

AGCATTTGGT GGGAACGTGA TGCCAAAGAT CTCTTGCCAG AAGGATTTAC TCATCCAGGT 720
TCACCAAACG GCGAGTTCAA AAAAGAAACT GATATCATGG ACGTTTGGTT TGACTCAGGT 780
TCATCATGGA ATGGAGTGGT GGTAAACCGT CCTGAATTGA CTTACCCAGC CGACCTTTAC 840
CTAGAAGGTT CTGACCAATA CCGTGGTTGG TTTAACTCAT CACTTATCAC ATCTGTTGCC 900
AACCATGGCG TAGCACCTTA CAAACAAATC TTGTCACAAG GTTTTGCCCT TGATGGTAAA 960
GGTGAGAAGA TGTCTAAATC TCTTGGAAAT ACCATTGCTC CAAGCGATGT TGAAAAACAA 1020
TTCGGTGCTG AAATCTTGCG TCTCTGGGTA ACAAGTGTTG ACTCAAGCAA TGACGTGCGT 1080
ATCTCTATGG ATATTTTGAG CCAAGTTTCT GAAACTTACC GTAAGATTCG TAACACTCTT 1140
CGTTTCTTGA TTGCCAATAC ATCTGACTTT AACCCAGCTC AAGATACAGT CGCTTACGAT 1200
GAGCTTCGTT CAGTTGATAA GTACATGACG ATTCGCTTTA ACCAGCTTGT CAAGACCATT 1260
CGTGATGCCT ATGCAGACTT TGAATTCTTG ACGATCTACA AGGCCTTGGT GAACTTTATC 1320
AACGTTGACT TGTCAGCCTT CTACCTTGAT TTTGCCAAAG ATGTTGTTTA CATTGAAGGT 1380
GCCAAATCAC TGGAACGCCG TCAAATGCAG ACTGTCTTCT ATGACATTCT TGTCAAAATC 1440
ACCAAACTCT TGACACCAAT CCTTCCTCAC ACTGCGGAAG AAATTTGGTC ATATCTTGAG 1500
TTTGAAACAG AAGACTTCGT CCAATTGTCA GAATTACCAG AGGCTCAAAC TTTTGCTAAT 1560
CAAGAAGAAA TCTTGGATAC ATGGGCAGCC TTCATGGACT TCCGTGGACA AGCTCAAAAA 1620
GCCTTGGAAG AAGCTCGTAA TGCAAAAGTA ATCGGTAAAT CACTTGAAGC ACACTTGACA 1680
GTTTATCCAA ACGAAGTTGT GAAAACTCTA CTCGAAGCAG TAAACAGCAA TGTGGCTCAA 1740
CTTTTGATCG TGTCAGACTT GACCATCGCA GAAGGACCAG CTCCAGAAGC TGCCCTTAGC 1800
TTCGAAGATG TAGCCTTCAC AGTTGAACGC GCTGCAGGTG AAGTATGTGA CCGTTGCCGT 1860
CGTATTGACC CAACAACAGC AGAACGTAGC TACCAGGCAG TTATCTGTGA CCACTGTGCA 1920
AGCATCGTAG AAGAAAACTT TGCGGAAGCA GTCGCAGAAG GATTTGAAGA GAAATAA 1977

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 658 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Leu Ser Glu Lys Phe Gly Trp Ala Asp Val Gln Val Leu Glu Thr Tyr
1 5 10 15
Arg Gly Gln Glu Leu Asn His Ile Val Thr Glu His Pro Trp Asp Thr
20 25 30
Ala Val Glu Glu Leu Val Ile Leu Gly Asp His Val Thr Thr Asp Ser
35 40 45
Gly Thr Gly Ile Val His Thr Ala Pro Gly Phe Gly Glu Asp Asp Tyr
50 55 60

Asn Val Gly Ile Ala Asn Asn Leu Glu Val Ala Val Thr Val Asp Glu
65 70 75 80

Arg Gly Ile Met Met Lys Asn Ala Gly Pro Glu Phe Glu Gly Gln Phe
85 90 95

Tyr Glu Lys Val Val Pro Thr Val Ile Glu Lys Leu Gly Asn Leu Leu
100 105 110

Leu Ala Gln Glu Glu Ile Ser His Ser Tyr Pro Phe Asp Trp Arg Thr
115 120 125

Lys Lys Pro Ile Ile Trp Arg Ala Val Pro Gln Trp Phe Ala Ser Val
130 135 140

Ser Lys Phe Arg Gln Glu Ile Leu Asp Glu Ile Glu Lys Val Lys Phe
145 150 155 160

His Ser Glu Trp Gly Lys Val Arg Leu Tyr Asn Met Ile Arg Asp Arg
165 170 175

Gly Asp Trp Val Ile Ser Arg Gln Arg Ala Trp Gly Val Pro Leu Pro
180 185 190

Ile Phe Tyr Ala Glu Asp Gly Thr Ala Ile Met Val Ala Glu Thr Ile
195 200 205

Glu His Val Ala Gln Leu Phe Glu Glu His Gly Ser Ser Ile Trp Trp
210 215 220

Glu Arg Asp Ala Lys Asp Leu Leu Pro Glu Gly Phe Thr His Pro Gly
225 230 235 240

Ser Pro Asn Gly Glu Phe Lys Lys Glu Thr Asp Ile Met Asp Val Trp
245 250 255

Phe Asp Ser Gly Ser Ser Trp Asn Gly Val Val Val Asn Arg Pro Glu
260 265 270

Leu Thr Tyr Pro Ala Asp Leu Tyr Leu Glu Gly Ser Asp Gln Tyr Arg
275 280 285

Gly Trp Phe Asn Ser Ser Leu Ile Thr Ser Val Ala Asn His Gly Val
290 295 300

Ala Pro Tyr Lys Gln Ile Leu Ser Gln Gly Phe Ala Leu Asp Gly Lys
305 310 315 320

Gly Glu Lys Met Ser Lys Ser Leu Gly Asn Thr Ile Ala Pro Ser Asp
325 330 335

Val Glu Lys Gln Phe Gly Ala Glu Ile Leu Arg Leu Trp Val Thr Ser
340 345 350

Val Asp Ser Ser Asn Asp Val Arg Ile Ser Met Asp Ile Leu Ser Gln
355 360 365

Val Ser Glu Thr Tyr Arg Lys Ile Arg Asn Thr Leu Arg Phe Leu Ile
370 375 380

Ala Asn Thr Ser Asp Phe Asn Pro Ala Gln Asp Thr Val Ala Tyr Asp
385 390 395 400

Glu Leu Arg Ser Val Asp Lys Tyr Met Thr Ile Arg Phe Asn Gln Leu
405 410 415

```
Val Lys Thr Ile Arg Asp Ala Tyr Ala Asp Phe Glu Phe Leu Thr Ile
            420                 425                 430
Tyr Lys Ala Leu Val Asn Phe Ile Asn Val Asp Leu Ser Ala Phe Tyr
        435                 440                 445
Leu Asp Phe Ala Lys Asp Val Val Tyr Ile Glu Gly Ala Lys Ser Leu
    450                 455                 460
Glu Arg Arg Gln Met Gln Thr Val Phe Tyr Asp Ile Leu Val Lys Ile
465                 470                 475                 480
Thr Lys Leu Leu Thr Pro Ile Leu Pro His Thr Ala Glu Glu Ile Trp
                485                 490                 495
Ser Tyr Leu Glu Phe Glu Thr Glu Asp Phe Val Gln Leu Ser Glu Leu
            500                 505                 510
Pro Glu Ala Gln Thr Phe Ala Asn Gln Glu Glu Ile Leu Asp Thr Trp
        515                 520                 525
Ala Ala Phe Met Asp Phe Arg Gly Gln Ala Gln Lys Ala Leu Glu Glu
    530                 535                 540
Ala Arg Asn Ala Lys Val Ile Gly Lys Ser Leu Glu Ala His Leu Thr
545                 550                 555                 560
Val Tyr Pro Asn Glu Val Val Lys Thr Leu Leu Glu Ala Val Asn Ser
                565                 570                 575
Asn Val Ala Gln Leu Leu Ile Val Ser Asp Leu Thr Ile Ala Glu Gly
            580                 585                 590
Pro Ala Pro Glu Ala Ala Leu Ser Phe Glu Asp Val Ala Phe Thr Val
        595                 600                 605
Glu Arg Ala Ala Gly Glu Val Cys Asp Arg Cys Arg Arg Ile Asp Pro
    610                 615                 620
Thr Thr Ala Glu Arg Ser Tyr Gln Ala Val Ile Cys Asp His Cys Ala
625                 630                 635                 640
Ser Ile Val Glu Glu Asn Phe Ala Glu Ala Val Ala Glu Gly Phe Glu
                645                 650                 655
Glu Lys
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CAACTTTTTG AAGAACATGG TTCAAGCATT TGGTGGGAAC GTGATGCCAA AGATCTCTTG    60
CCAGAAGGAT TTACTCATCC AGGTTCACCA AACGGCGAGT TCAAAAAAGA AACTGATATC   120
ATGGACGTTT GGTTTGACTC AGGTTCATCA TGGAATGGAG TGGTGGTAAA CCGTCCTGAA   180
TTGACTTACC CAGCCGACCT TTACCTAGAA GGTTCTGACC AATACCGTGG TTGGTTTAAC   240
TCATCACTTA TCACATCTGT TGCCAACCAT GGCGTAGCAC CTTACAAACA AATCTTGTCA   300
CAAGGTTTTG CCCTTGATGG TAAAGGTGAG AAGATGTCTA AATCTCTTGG AAATACCATT   360
GCTCCAAGCG ATGTTGAAAA ACAATTCGGG                                   390
```

What is claimed is:

1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 6;

(b) a polynucleotide which is complementary to the polynucleotide of (a);

(c) a polynucleotide having at least a 70% identity to a polynucleotide encoding the same mature polypeptide expressed by the ileS gene contained in the *Streptococcus pneumoniae* of the deposited strain; and

(d) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b) or (c).

2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.

3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.

4. The polynucleotide of Claim 2 comprising the nucleic acid sequence set forth in SEQ ID NO:1, 5 or 7.

5. The polynucleotide of Claim 2 comprising the polynucleotide sequence set forth in SEQ ID NO:1, 5 or 7.

6. The polynucleotide of Claim 2 which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 6.

7. A vector comprising the polynucleotide of Claim 1.

8. A host cell comprising the vector of Claim 7.

9. A process for producing a polypeptide comprising: expressing from the host cell of Claim 8 a polypeptide encoded by said DNA.

10. A process for producing a ileS polypeptide or fragment comprising culturing a host of claim 8 under conditions sufficient for the production of said polypeptide or fragment.

11. A polypeptide comprising an amino acid sequence which is at least 70% identical to the amino acid sequence of SEQ ID NO:2 or 6.

12. A polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:2 or 6.

13. An antibody against the polypeptide of claim 11.

14. An antagonist which inhibits the activity or expression of the polypeptide of claim 11.

15. A method for the treatment of an individual in need of ileS polypeptide comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 11.

16. A method for the treatment of an individual having need to inhibit ileS polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 14.

17. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 11 in an individual comprising:

(a) determining a nucleic acid sequence encoding said polypeptide, and/or

(b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.

18. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 11 comprising:

contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;

and determining whether the compound interacts with and activates or inhibits an activity of the polypetide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.

19. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with ileS polypeptide of claim 11, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said animal from disease.

20. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of ileS polypeptide of claim 11, or fragment or a variant thereof, for expressing said ileS polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/06551

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/04; C12N 15/00, 1/20, 9/00; A61K 38/43
US CL :536/23.2, 24.3; 435/320.1, 252.3, 183; 424/94.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.2, 24.3; 435/320.1, 252.3, 183; 424/94.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	CHALKER et al. Analysis and Toxic Overexpression in Escherichia coli of a Staphylococcal Gene Encoding Isoleucyl-tRNA Synthetase. Gene. April 1994. Vol. 141. No. 1. pages 103-108, see entire document.	1-3, 7, 8, 10 --- 1-12, 15
X --- Y	CSANK et al. Isoleucyl-tRNA Synthetase From the Ciliated Protozoan Tetrahymena thermophila. J. Biol. Chem. March 1992. Vol. 267. No. 7. pages 4592-4599, see entire document.	1-3, 7 --- 1-12, 15
Y	VON DEN HAAR et al. Target Directed Drug Synthesis: The Aminoacyl-tRNA Synthetases as Possible Targets. Angewandte Chemie. March 1981. Vol. 20. No. 3. pages 217-302, see entire document.	1-12, 15

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 JULY 1997	03 SEP 1997

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	LISA J. HOBBS, PH.D.
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MEINNEL et al. 'Aminoacyl-tRNA Synthetases: Occurence, Structure, and Function." In: tRNA: Structure, Biosynthesis, and Function. Edited by D. Soll and U. RajBhandary. Washington, DC: American Society for Microbiology. 1985. pages 251-292, see entire document.	1-12, 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/06551

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-12 and 15

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS and STN (bioscience and patents indexes): Streptococcus pneumoniae, S.pneumoniae, Streptococcus, tRNA synthetase#, tRNA ligase#, transfer RNA synthetase# and transfer RNA ligase#. GenBank, embl, N-Geneseq, EST, A-Geneseq, PIR, Swissprot: Seq. ID Nos. 1, 2, 5, 6 and 7.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I. Claims 1-12 and 15, drawn to DNA molecules, recombinant methods of production of the protein, the protein product and a method of treatment using the protein.
Group II. Claim 13, drawn to an antibody.
Group III. Claim 14, drawn to an antagonist.
Group IV. Claim 16, drawn to a method of treatment using the antagonist.
Group V. Claim 17, drawn to a process for disease diagnosis.
Group VI. Claim 18, drawn to a method of identification of inhibitors and effectors.
Group VII. Claim 19, drawn to a method of producing the antibody.
Group VIII. Claim 20, drawn to gene therapy.

The inventions listed as Groups I-VIII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I shares the special technical feature of the recombinant method of production of the protein, which the other groups do not share. Group II shares the special technical feature of the antibody, which the other groups do not share. Group III shares the special technical feature of the antagonist, which the other groups do not share. Group IV shares the special technical feature of a method of treatment, which the other groups do not share. Group V shares the special technical feature of the disease diagnosis process, which the other groups do not share. Group VI shares the special technical feature of the method of identification of inhibitors and effectors, which the other groups do not share. Group VII shares the special technical feature of a method for the production of an antibody, which the other groups do not share. Group VIII shares the special technical feature of gene therapy, which the other groups do not share.